                 Real-time Transport Protocol (RTP)
             Payload Format for Enhanced AC-3 (E-AC-3) Audio

Status of This Memo

   This document specifies an Internet standards track protocol for the
   Internet community, and requests discussion and suggestions for
   improvements. Please refer to the current edition of the "Internet
   Official Protocol Standards" (STD 1) for the standardization state
   and status of this protocol. Distribution of this memo is unlimited.

Abstract

   This document describes a Real-time Transport Protocol (RTP) payload
   format for transporting Enhanced AC-3 (E-AC-3) encoded audio data.
   E-AC-3 is a high-quality, multichannel audio coding format and is an
   extension of the AC-3 audio coding format, which is used in US High-
   Definition Television (HDTV), DVD, cable and satellite television,
   and other media. E-AC-3 is an optional audio format in US and world
   wide digital television and high-definition DVD formats. The RTP
   payload format as presented in this document includes support for
   data fragmentation.

Table of Contents

1. Introduction

   The Enhanced AC-3 (E-AC-3) [ETSI] audio coding system is built on a
   foundation of AC-3. It is an enhancement and extension to AC-3,
   which is an existing audio coding standard commonly used for DVD,
   broadcast, cable, and satellite television content. E-AC-3 is
   designed to enable operation at both higher and lower data rates than
   AC-3, provide expanded channel configurations, and provide greater
   flexibility for carriage of multiple audio program elements. The
   relationship between E-AC-3 and AC-3 provides for low-loss, low-cost
   conversion between the two and makes E-AC-3 especially suitable in
   applications that require compatibility with the existing broadcast-
   reception and audio/video decoding infrastructure. Dolby Digital
   Plus is a branded version of Enhanced AC-3.

   E-AC-3 has been standardized within both the European
   Telecommunications Standards Institute (ETSI) and the Advanced
   Television Systems Committee (ATSC). It is an optional audio format
   for use in US (ATSC) and Digital Video Broadcasting (DVB) television
   transmission. It is also a required audio format for use in the High
   Definition (HD)-DVD optical-storage media format and included in the
   Blu-ray Disc format.

There is a need to stream E-AC-3 content over IP networks. E-AC-3 is
primarily used in audio-for-video applications, so RTP serves well as
a transport solution with its mechanism for synchronizing streams.
Applications for streaming E-AC-3 include Internet Protocol
television (IPTV), video on demand, interactive features of next
generation DVD formats, and transfer of movies across a home network.

Section 2 gives a brief overview of the E-AC-3 algorithm. Section 3
specifies values for fields in the RTP header, and Section 4
specifies the E-AC-3 payload format, itself. Section 5 discusses
media types and Session Description Protocol (SDP) usage. Security
considerations are covered in Section 6, congestion control in
Section 7, and IANA considerations in Section 8.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in [RFC2119].

2. Overview of Enhanced-AC-3

Enhanced AC-3 (E-AC-3) is a frequency-domain perceptual audio coding
system. Time blocks of an audio signal are converted from the time
domain to the frequency domain by a transform (the Modified Discrete
Cosine Transform (MDCT)) so that a model of the human auditory
perceptual system can be applied. In this domain, quantization noise
can be constrained to specific frequency regions. The perceptual
model predicts in which frequency regions the auditory system will be
least able to detect the quantization noise from data rate reduction.
A more detailed technical description of E-AC-3 can be found in
[2004AES].

E-AC-3 is built upon a foundation of AC-3. More background on AC-3
can be found in the AC-3 specification [ETSI], a technical paper
[1994AES], and the AC-3 RTP payload format [RFC4184]. The frame
structure and meta-data of AC-3 are maintained. E-AC-3 content is
not directly compatible with AC-3 decoders, but it can be converted
to the AC-3 format to provide compatibility with existing decoders.
Because AC-3 is the foundation of E-AC-3, conversion between the two
formats can be done in a way that minimizes the degradations
associated with tandem coding. In addition, the computational cost
of the conversion is reduced compared to a full decode and re-encode.

E-AC-3 exploits psychoacoustic phenomena that cause a significant
fraction of the information contained in a typical audio signal to be
inaudible. Substantial data reduction occurs via the removal of
inaudible information contained in an audio stream. Source coding
techniques are further used to reduce the data rate.

Like most perceptual coders, E-AC-3 operates in the frequency domain.
A 512-point MDCT transform is taken with 50% overlap, providing 256
new frequency samples. Frequency samples are then converted to
exponents and mantissas. Exponents are differentially encoded.
Mantissas are allocated a varying number of bits depending on the
audibility of the spectral components associated with them.
Audibility is determined via a masking curve. Bits for mantissas are
allocated from a global bit pool.

E-AC-3 adds new coding tools, such as a longer filter bank, vector
quantization, and spectral extension, to provide greater data
efficiency and to operate at lower data rates than AC-3. In the
other direction, an expanded bit stream syntax and new frame
constraints permit operation at higher data rates than AC-3. The
E-AC-3 syntax also allows a larger number of audio channels in one
bit stream. E-AC-3 operates at data rates from 32 kbps to 6.144 Mbps
and at three sampling rates: 32 kHz, 44.1 kHz, and 48 kHz.

E-AC-3 supports the carriage of multiple programs and the carriage of
programs with more than a baseline of 5.1 audio channels. Both of
these extensions beyond AC-3 are accomplished by time multiplexing
additional data with baseline data. In the case of multiple
programs, frames with data for the programs are interleaved. In the
case of more than 5.1 channels, frames from substreams carrying the
extra channels are interleaved with the independent substream that
carries a 5.1-channel compatible mix. Both of these forms of
multiplexing can occur in the same bit stream. In other words,
mixing multiple programs, some or all with more than 5.1 channels, is
permitted.

Additional channel capacity is enabled by adding substreams to a
program. One primary substream, called the "independent substream",
is required for each program. This substream carries a self-
contained mix of the audio, using a maximum of 5.1 channels, which
makes its channel configuration compatible with AC-3. Then,
additional, optional substreams are used in the program to carry
additional channels. The data for each additional channel carries an
indication of whether that channel provides data for an additional
speaker location or replacement data for one of the speaker locations
already defined by a previous substream. For example, one common
7.1-channel format uses three front channels and four surround
channels. It is packaged with a primary substream, which contains a
5.1-channel downmix of the 7.1-channel content, using left, center,
right, left surround, right surround, and low-frequency effects
channels. One dependent substream supplies four channels:
replacements for left surround and right surround, along with two
additional surround channels (left back and right back).

The specification for E-AC-3 [ETSI] requires that all E-AC-3 decoders
be capable of decoding at least a baseline portion of any E-AC-3 bit
stream, which consists of the first independent substream of the
first program, and of ignoring the other elements of the bit stream.
This baseline is limited to 5.1 channels, and a system is also able
to convert to configurations with fewer channels for a presentation
that matches its output capabilities, if needed. More capable
decoders can optionally choose among and mix multiple programs, and
also decode configurations with more channels than the baseline by
decoding dependent substreams.

2.1. E-AC-3 Bit Stream

2.1.1. Sync Frames and Audio Blocks

The basic organizational building block in an E-AC-3 bit stream is
the sync frame (also called a frame in this document). A sync frame
contains the data necessary to decode time domain audio samples for
one or more channels over a time of one or more audio blocks, so a
frame is an Application Data Unit (ADU). Each E-AC-3 frame contains
a Sync Information (SI) field, a Bit Stream Information (BSI) field,
an Audio Frame (AF) field, and up to six audio blocks (ABs). Each AB
represents 256 Pulse Code Modulation (PCM) samples for each channel.
The frame ends with an optional auxiliary data field (AUX) and an
error correction field (CRC). Figure 1 shows the structure of an
E-AC-3 frame, where N is the number of blocks in the frame.

```
    +---+---+---+---------+- ... -+---------+---+---+
    |SI |BSI|AF |  AB(0)  |  ...  |  AB(N)  |AUX|CRC|
    +---+---+---+---------+- ... -+---------+---+---+
```
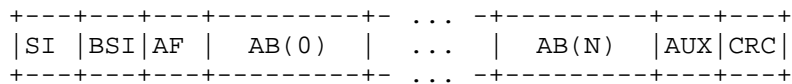
        Figure 1. E-AC-3 frame format with more than one block

The SI field contains information needed to acquire and maintain
codec synchronization. The BSI field contains parameters that
describe the coded audio service. It carries an indication of the
size of the frame in 16-bit words ('frmsiz', Section E.1.3 of [ETSI])
and an indication of the sampling rate ('fscod'). It also carries an
indication of the number of blocks in the frame ('numblkscod');
permitted values are one, two, three, or six blocks. The AF field
contains information about coding tools that applies to the entire
frame. Each block has a duration of 256 samples, so a frame's
duration is the corresponding multiple of 256 samples. The time
duration of the frame is also dependent on the sampling rate, as
shown in Table 1.

Table 1. Time duration of E-AC-3 frame (number of blocks vs.
sampling rate)

```
+------------------+--------+----------------+----------------+
| blocks per frame | 32 kHz |       44.1 kHz |         48 kHz |
+------------------+--------+----------------+----------------+
|                1 |   8 ms | approx. 5.8 ms | approx. 5.3 ms |
|                2 |  16 ms | approx. 11.6 ms| approx. 10.7 ms|
|                3 |  24 ms | approx. 17.4 ms|          16 ms |
|                6 |  48 ms | approx. 34.8 ms|          32 ms |
+------------------+--------+----------------+----------------+
```

Each audio block contains header fields that indicate the use of
various coding tools: block switching, dither, coupling, spectral
extension, and exponent strategy. They also contain metadata,
optionally used to enhance playback, such as dynamic range control.
Finally, the exponents and bit allocation data needed to decode the
mantissas into audio data, and the mantissas themselves, are
included. The format of audio blocks is described in detail in
[ETSI].

2.1.2. Programs and Substreams

An E-AC-3 bit stream is logically arranged into programs. A bit
stream contains one or more programs, up to a maximum of eight. When
multiple programs are present in a bit stream, the frames that
constitute them are interleaved in time.

```
   +----------+-     -+----------+----------+-     -+----------+-
   |Program(1)|  ...  |Program(N)|Program(1)|  ...  |Program(N)| ...
   | Frame 0  |       | Frame 0  | Frame 1  |       | Frame 1  |
   +----------+-     -+----------+----------+-     -+----------+-
```
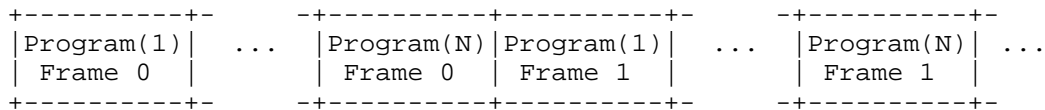
Figure 2. Interleaving of multiple programs in an E-AC-3 bit stream

Each program contains one independent substream and optionally
contains up to eight dependent substreams. The independent substream
carries a soundtrack of up to 5.1 channels, the multichannel format
that matches the capabilities of AC-3, and can be meaningfully
decoded and presented without any of the associated dependent
substreams. The dependent substreams are used to provide alternate
channel data that enable different channel configurations, for
example, to increase the number of channels beyond 5.1. A frame of a
dependent substream can be decoded by itself, but its content can
only be meaningfully presented in conjunction with the corresponding
independent substream. The type and identity of the substream to
which a frame belongs can be determined from parameters in the
frame's BSI (strmtyp and substreamid, in Section E.1.3.1 of [ETSI]).

When a program contains more than one substream, the frames belonging
to those substreams are interleaved in time, and taken together, the
frames of a program that correspond to the same time period are
called a 'program set'. Figure 3 shows the interleaving of
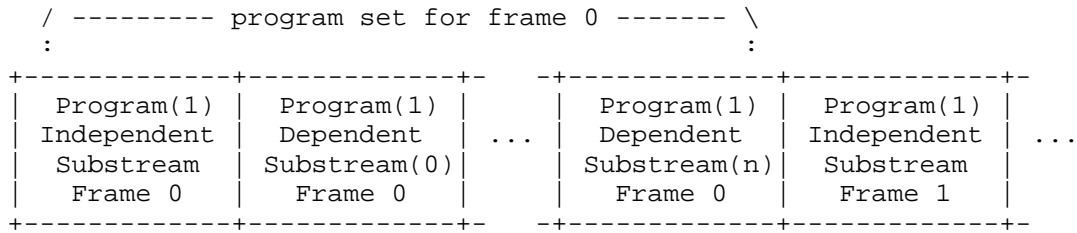substreams for a single program.

```
   / --------- program set for frame 0 ------- \
   :                                            :
+-------------+-------------+-   -+-------------+-------------+-
|  Program(1) |  Program(1) |     |  Program(1) |  Program(1) |
| Independent |  Dependent  | ... |  Dependent  | Independent | ...
|  Substream  | Substream(0)|     | Substream(n)|  Substream  |
|   Frame 0   |   Frame 0   |     |   Frame 0   |   Frame 1   |
+-------------+-------------+-   -+-------------+-------------+-
```

   Figure 3. Interleaving of multiple substreams in an E-AC-3 program

2.1.3. Frame Sets

A further logical organization of the E-AC-3 bit stream is applied to
facilitate conversion of E-AC-3 bit streams to AC-3 bit streams. In
this organization, the frames carrying six consecutive audio blocks
are treated as a group, called a 'frame set', regardless of the
number of frames needed to carry six audio blocks. This grouping
extends across all programs and substreams that cover the time period
of the six blocks. Since E-AC-3 frames may carry one, two, three, or
six blocks, a frame set will consist of six, three, two, or one
frames. AC-3 frames always carry six blocks, so the frame set
provides framing synchronization between an E-AC-3 bit stream and an
AC-3 bit stream. Metadata that indicates the alignment is carried in
the first frame (which will be part of an independent substream) of
each frame set in an E-AC-3 stream. This first frame can be
identified by a parameter in the BSI field of the bit stream: the
Converter Synchronization flag (convsync, in Section E.1.3.1.34 of
[ETSI]) is set to true (1).

3. RTP E-AC-3 Header Fields

The RTP header is defined in the RTP specification [RFC3550]. This
section defines how a number of fields in the header are used.

o Payload Type (PT): The assignment of an RTP payload type for this
   packet format is outside the scope of this document; it is
   specified by the RTP profile under which this payload format is
   used, or signaled dynamically out-of-band (e.g., using SDP).

o Marker (M) bit: The M bit is set to one to indicate that the RTP
   packet payload contains at least one complete E-AC-3 frame or
   contains the final fragment of an E-AC-3 frame.

o Extension (X) bit: Defined by the RTP profile used.

o Timestamp: A 32-bit word that corresponds to the sampling instant
   for the first E-AC-3 frame in the RTP packet. Packets containing
   fragments of the same frame MUST have the same timestamp. The
   timestamp of the first RTP packet sent SHOULD be selected at
   random; thereafter, it increases linearly according to the number
   of samples included in each frame. Note that the number of
   samples in a frame depends on the number of blocks in the frame,
   with 256 samples in each block. Also note that more than one
   frame might correspond to the same time period when multiple
   channel configurations or programs are present. If these frames
   occupy multiple packets, it is possible that the resulting packets
   will have the same timestamp value.

4. RTP E-AC-3 Payload Format

   This payload format is defined for E-AC-3, as defined in Annex E of
   [ETSI]. Note that E-AC-3 decoders are required to be capable of
   decoding AC-3 bit streams, so a receiver capable of receiving the
   E-AC-3 payload format defined in this document MUST also receive the
   payload format for AC-3 defined in [RFC4184].

   According to [RFC2736], RTP payload formats should contain an
   integral number of application data units (ADUs). The E-AC-3 frame
   corresponds to an ADU in the context of this payload format. Each
   RTP payload MUST start with the two-byte payload specific header
   followed by an integral number of complete E-AC-3 frames, or a single
   fragment of an E-AC-3 frame.

   If an E-AC-3 frame exceeds the MTU for a network, it SHOULD be
   fragmented for transmission within an RTP packet. Section 4.2
   provides guidelines for creating frame fragments.

4.1. Payload Specific Header

   There is a two-octet Payload header at the beginning of each payload.
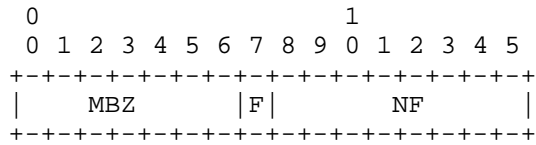   Each E-AC-3 RTP payload MUST begin with the following Payload header.

```
         0                   1
         0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |    MBZ      |F|      NF       |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

             Figure 4. E-AC-3 RTP Payload header

   o Must Be Zero (MBZ): Bits marked MBZ SHALL be set to the value zero
      and SHALL be ignored by receivers. The bits are reserved for
      future extensions.

   o Frame Type (F): This one-bit field indicates the type of frame(s)
      present in the payload. It takes the following values: 0 - One
      or more complete frames. 1 - Fragment of frame. (Note that the M
      bit in the RTP header is set for the final fragment.)

   o Number of frames/fragments (NF): An 8-bit field whose meaning
      depends on the Frame Type (F) in this payload. For complete
      frames (F of 0), it is used to indicate the number of E-AC-3
      frames in the RTP payload. For frame fragments (F of 1), it is
      used to indicate the number of fragments (and therefore packets)
      that make up the current frame. NF MUST be identical for packets
      containing fragments of the same frame.

   When receiving E-AC-3 payloads with F = 0 and more than a single
   frame (NF > 1), a receiver needs to use the "frmsiz" field in the BSI
   header in each E-AC-3 frame to determine the frame's length if the
   receiver needs to determine the boundary of the next frame. Note
   that the frame length varies from frame to frame in some
   circumstances.

4.2. Fragmentation of E-AC-3 Frames

   The size of an E-AC-3 frame is signaled in the Frame Size (frmsiz)
   field in a frame's BSI header. The value of this field is one less
   than the number of 16-bit words in the frame. If the size of an
   E-AC-3 frame exceeds the MTU size, the frame SHOULD be fragmented at
   the RTP level. The fragmentation MAY be performed at any byte
   boundary in the frame. RTP packets containing fragments of the same
   E-AC-3 frame SHALL be sent in consecutive order, from first to last
   fragment. This enables a receiver to assemble the fragments in the
   correct order.

4.3. Concatenation of E-AC-3 Frames

   There are cases where E-AC-3 frame sizes are smaller than the MTU
   size and it is advantageous to include multiple frames in a packet.

It is useful to take into account the logical arrangement of the bit
stream into program sets and frame sets to constrain the effects of
the loss of a packet. It is desirable for a complete program set or
a complete frame set to be included in one packet. Also, it is
undesirable for frames from more than one program set or frame set to
be in the same packet, unless the sets are complete. In this way,
the loss of a packet is kept from causing the contents of another
packet to be unusable.

Frames from more than one program set SHOULD NOT be included in the
same packet unless all program sets in the packet are complete.
Frames from more than one frame set SHOULD NOT be included in the
same packet unless all frame sets in the packet are complete.

4.4. Carriage of AC-3 Frames

The E-AC-3 specification [ETSI] requires that E-AC-3 decoders be
capable of decoding AC-3 frames. That specification also supports
carriage of AC-3 frames in an E-AC-3 bit stream. Due to differences
between E-AC-3 and AC-3 frames, there are restrictions placed on the
use of AC-3 frames: they are only used for the independent substream
of the first (or only) program in an E-AC-3 bit stream. Note that
carriage of only E-AC-3 frames, only AC-3 frames, and a mixture of
E-AC-3 and AC-3 frames are all legal configurations. It is legal to
change among the configurations in a bit stream. The AC-3 frame
format is described in [RFC4184] and specified in [ETSI].

5. Types and Names

5.1. Media Type Registration

This registration uses the template defined in [RFC4288] and follows
[RFC3555].

To: ietf-types@iana.org
Subject: Registration of media type audio/eac3

Type name: audio

Subtype name: eac3

Required parameter:

o   rate: The RTP timestamp clock rate that is equal to the audio
    sampling rate. Permitted rates are 32000, 44100, and 48000.

   Optional parameter:

   o bitStreamConfig: The configuration of programs and substreams in
      the bit stream, expressed as a sequence of ASCII characters. This
      parameter can serve two purposes. First, during the creation of a
      session, the bitStreamConfig parameter might be used to negotiate
      a match between the requirements of a bit stream and the
      capabilities of a receiver to avoid using network bandwidth for
      data that cannot be used. Second, it makes the configuration of
      the bit stream explicit to the receiver so that whenever a packet
      is lost, the receiver can identify which kind of frame(s) has been
      lost to aid error mitigation.

      The format for the value for this parameter is to represent each
      substream of the bit stream by a single character indicating its
      type, immediately followed by the number of audio channels
      resulting if a frame of that substream (plus any other required
      substreams) is decoded. Note that even though Low-Frequency
      Effects (LFE) channels are often described as "fractional"
      channels (e.g., the ".1" in 5.1), for this parameter, an LFE
      channel is counted as one (e.g., a 5.1-channel configuration is
      indicated as 6). The configuration of the bit stream MUST match
      the value of this parameter for the duration of the session.

      Allowed values for the substream type are as follows:

      i - Independent substream.
      d - Dependent substream.

   The E-AC-3 specification [ETSI] defines which configurations of bit
   streams are legal, which constrains the values the bitStreamConfig
   parameter will take. Each program starts with, and contains exactly
   one, independent substream ('i'). Each independent substream is
   followed by between 0 and 8 dependent substreams ('d'), which belong
   to the same program. See Section 2.1.2 for more discussion of
   programs and substreams.

   For example, consider a bit stream containing two programs:

   * the first program with

      + a six-channel independent substream
      + a dependent substream containing the additional channels needed
         for eight channels
      + a second dependent substream containing the further channels
         needed for 14 channels

   * along with a second program with

      + another six-channel independent substream
      + a dependent substream containing the additional channels needed
         for eight channels

   Then the configuration of the bit stream is indicated as follows:

      bitStreamConfig = i6d8d14i6d8

   When the bitStreamConfig parameter is being used in an offer/answer
   exchange, zero (0) for the number of channels for a substream in an
   answer is used to indicate a substream that the answerer desires not
   to receive.

   Encoding considerations:

      This media type is framed and contains binary data.

   Security considerations:

      See Section 6 of RFC 4598.

   Interoperability considerations:

   To maintain interoperability with AC-3-capable end-points, in cases
   where negotiation is possible, an E-AC-3 end-point SHOULD declare
   itself also as AC-3 capable (i.e., supporting also "audio/ac3" as
   specified in RFC 4184 [RFC4184]). Note that all E-AC-3 end-points
   are required to be AC-3 capable.

   Published specification:

      RFC 4598 and ETSI TS 102.366 [ETSI].

   Applications that use this media type:

      Multichannel audio compression of audio, and audio for video.

   Additional information:

      Magic number(s): The first two octets of an E-AC-3 frame are
         always the synchronization word, which has the hex value
         0x0B77.

   Person & email address to contact for further information:

      Brian Link <bdl@dolby.com> IETF AVT working group.

   Intended usage:

      COMMON

   Restrictions on usage:

      This media type depends on RTP framing, and hence is only defined
      for transfer via RTP [RFC3550]. Transport within other framing
      protocols is not defined at this time.

   Author/Change controller:

      IETF Audio/Video Transport Working Group delegated from the IESG.

5.2. SDP Usage

   The information carried in the media type specification has a
   specific mapping to fields in the Session Description Protocol (SDP)
   [RFC2327], which is commonly used to describe RTP sessions. When SDP
   is used to specify sessions employing E-AC-3, the mapping is as
   follows:

   o The Media type ("audio") goes in SDP "m=" as the media name.

   o The Media subtype ("eac3") goes in SDP "a=rtpmap" as the encoding
      name.

   o The required parameter "rate" also goes in "a=rtpmap" as the clock
      rate. (The optional "channels" rtpmap encoding parameter is not
      used. Instead, the information is included in the optional
      parameter bitStreamConfig.)

   o The optional parameter "bitStreamConfig" goes in the SDP "a=fmtp"
      attribute.

   The following is an example of the SDP data for E-AC-3:

        m=audio 49111 RTP/AVP 100
        a=rtpmap:100 eac3/48000
        a=fmtp:100 bitStreamConfig i6d8d14i6d8

   Certain considerations are needed when SDP is used to perform
   offer/answer exchanges [RFC3264].

   o The "rate" is a symmetric parameter, and the answer MUST use the
      same value or the answerer removes the payload type.

o The "bitStreamConfig" parameter is declarative and indicates, for
   sendonly, the intended arrangement of substreams in the bit
   stream, along with the channel configuration, to transmit, and for
   recvonly or sendrecv, the desired bit stream arrangement and
   channel configuration to receive. The format of the
   bitStreamConfig value in an answer MAY differ from the offer value
   by replacing the number of channels for any undesired substreams
   with '0'. It is valid to zero out dependent substreams containing
   undesired channel configurations and to zero out all the
   substreams of an undesired program. Then the sender MAY reoffer
   the stream in the receiver's preferred configuration if it is
   capable of providing that configuration. Note that all receivers
   are capable of receiving, and all decoders are capable of
   decoding, any of the legal bit stream configurations, so the
   parameter exchange is not needed for interoperability. The
   parameter exchange might be used to help optimize the transmission
   to the number of programs or channels the receiver requests.

o Since an AC-3 bit stream is a special case of an E-AC-3 bit
   stream, it is permissible for an AC-3 bit stream to be carried in
   the E-AC-3 payload format. To ensure interoperability with
   receivers that support the AC-3 payload format but not the E-AC-3
   payload format, a sender that desires to send an AC-3 bit stream
   in the E-AC-3 payload format SHOULD also offer the session in the
   AC-3 payload format by including payload types for both media
   subtypes: 'ac3' and 'eac3'.

6. Security Considerations

   The payload format described in this document is subject to the
   security considerations defined in RTP [RFC3550] and in any
   applicable RTP profile (e.g., [RFC3551]). To protect the user's
   privacy and any copyrighted material, confidentiality protection
   would have to be applied. To also protect against modification by
   intermediate entities and ensure the authenticity of the stream,
   integrity protection and authentication would be required.
   Confidentiality, integrity protection, and authentication have to be
   solved by a mechanism external to this payload format, for example,
   Secure Real-time Transport Protocol (SRTP) [RFC3711].

   The E-AC-3 format is designed so that the validity of data frames can
   be determined by decoders. The required decoder response to a
   malformed frame is to discard the malformed data and conceal the
   errors in the audio output until a valid frame is detected and
   decoded. This is expected to prevent crashes and other abnormal
   decoder behavior in response to errors or attacks.

7. Congestion Control

   The general congestion control considerations for transporting RTP
   data apply to E-AC-3 audio over RTP as well; see RTP [RFC3550], and
   any applicable RTP profile (e.g., [RFC3551]).

   E-AC-3 is a variable bit rate coding system so it is possible to use
   a variety of techniques to adapt to network bandwidth.

8. IANA Considerations

   The IANA has registered a new media subtype for E-AC-3 (see Section
   5).

9. References

9.1. Normative References

   [ETSI]   ETSI, "Digital Audio Compression (AC-3, Enhanced AC-3)
               Standard", TS 102 366, February 2005.

   [RFC2119]   Bradner, S., "Key words for use in RFCs to Indicate
               Requirement Levels", BCP 14, RFC 2119, March 1997.

   [RFC4184]   Link, B., Hager, T., and J. Flaks, "RTP Payload Format for
               AC-3 Audio", RFC 4184, October 2005.

   [RFC3550]   Schulzrinne, H., Casner, S., Frederick, R., and V.
               Jacobson, "RTP: A Transport Protocol for Real-Time
               Applications", STD 64, RFC 3550, July 2003.

   [RFC4288]   Freed, N. and J. Klensin, "Media Type Specifications and
               Registration Procedures", BCP 13, RFC 4288, December 2005.

   [RFC3555]   Casner, S. and P. Hoschka, "MIME Type Registration of RTP
               Payload Formats", RFC 3555, July 2003.

   [RFC2327]   Handley, M. and V. Jacobson, "SDP: Session Description
               Protocol", RFC 2327, April 1998.

   [RFC3264]   Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model
               with Session Description Protocol (SDP)", RFC 3264, June
               2002.

9.2. Informative References

   [2004AES] Fielder, L., Andersen, R., Crockett, B., Davidson, G.,
              Davis, M., Turner, S., Vinton, M., and P. Williams,
              "Introduction to Dolby Digital Plus, an Enhancement to the
              Dolby Digital Coding System", Preprint 6196, Presented at
              the 117th Convention of the Audio Engineering Society,
              October 2004.

   [1994AES] Todd, C., Davidson, G., Davis, M., Fielder, L., Link, B.,
              and S. Vernon, "AC-3: Flexible Perceptual Coding for Audio
              Transmission and Storage", Preprint 3796, Presented at the
              96th Convention of the Audio Engineering Society, May
              1994.

   [RFC2736] Handley, M. and C. Perkins, "Guidelines for Writers of RTP
              Payload Format Specifications", BCP 36, RFC 2736, December
              1999.

   [RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and
              Video Conferences with Minimal Control", STD 65, RFC 3551,
              July 2003.

   [RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K.
              Norrman, "The Secure Real-time Transport Protocol (SRTP)",
              RFC 3711, March 2004.

Author's Address

   Brian Link
   Dolby Laboratories
   100 Potrero Ave.
   San Francisco, CA 94103
   US

   Phone: +1 415 558 0200
   EMail: bdl@dolby.com

Full Copyright Statement

Intellectual Property

Acknowledgement